Mathematics for Data Science

Learning Outcomes

By the end of this session, you will have covered the following learning outcomes:

- **Demonstrate** knowledge of performing arithmetic operations on polynomials.
- **Perform** vector arithmetic operations such as addition, subtraction, multiplication, division, dot product, and multiplication with a scalar.
- **Perform** matrix operations such as addition, subtraction, multiplication and understand the intuition behind the process.
- **Demonstrate** knowledge of rates of change and using derivatives to analyze functions.
- **Explain** why probability is essential to statistics and data science.

Overview

Linear Algebra

We use linear algebra in data preprocessing, data transformation, and model evaluation. We should learn linear algebra because of the following reasons:

- We represent datasets in the form of a matrix. In contrast, we use vectors to define individual variables, such as the response variable/ target variable in machine learning.
- We define algorithms using vector and matrix notation. The understanding linear algebra will enable us to read descriptions of existing algorithms in textbooks or other resources.

Data science's important linear algebra concepts include vectors, matrices, matrix operations (transpose, inverse, determinant, and eigenvalues).

Probability

We use probability concepts to estimate the likelihood of an event occurring. For example, if we want to predict an outcome of a variable that can take one of many available values, we have to involve the mathematics of probability. A few probability concepts that we might use in data science projects include:

- We use **probability distributions** while collecting data. Datasets used in most cases represent a sample from a population. Using this sample, we can find distinctive patterns in the data that can help us make predictions about our main inquiry topic.
- **Distribution Characteristics:** The mean, the Variance, and the standard deviation tell us different things about the distribution's shape and behavior.

- **Conditional Probability:** Several algorithms depend on the Bayes theorem, a formula that demonstrates the probability of an event depending on prior knowledge about the conditions associated with the event. An example of such an algorithm is the Naive Bayes Algorithm.

Data science's important concepts include conditional probability and dependence, binomial variables and distributions, sampling, and sample distributions.

Calculus

Calculus is the branch of mathematics studying the rate of change quantities (or slopes of curves). A few uses of calculus in data science include:

- Optimization algorithms like gradient descent use derivatives to decide whether to increase or decrease weights to maximize or minimize some objective (e.g., a model's accuracy or error functions).
- We use calculus to understand how functions change over time (derivatives) and calculate the total quantity accumulated (integrals).

Important calculus concepts to learn include limits, differentiation, derivatives, and multivariate differentiation.

Practice

- **Practice Notebook:** Equations, Linear Equations, Factorization, Functions, Differentiation, and Probability. [https://bit.ly/MathsforDataScience]
  - This notebook will help you understand the importance of basic mathematics concepts in data science. After going through the above notebook, you should go through the following quiz. [https://bit.ly/MathsforDSQuiz]
- **Practice Notebook:** Linear Algebra Basics (Vectors and Matrices) [Link]

Project

- **Project Brief:** Mathematics for Data Science with Python [Link]